View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Hasanuddin University Repository

Estimasi Parameter Distribusi *Gumbel* menggunakan Metode Regresi *Rank* Kuadrat Terkecil Terboboti

Jumriana[*1], Anisa[2], Andi Kresna Jaya[3]

ABSTRAK

Metode kuadrat terkecil yang didasarkan pada hubungan antara fungsi distribusi kumulatif empiris dan statistik terurut disebut juga regresi *rank*, dapat digunakan untuk mengestimasi parameter dari distribusi. Pada penelitian ini, metode kuadrat terkecil terboboti digunakan untuk mengestimasi parameter dari distribusi *Gumbel*, dimana nilai bobot sebanding dengan invers dari variansi sampel yang besar dari sebuah fungsi dari statistik terurut. Metode ini digunakan sebagai alternatif dari metode kuadrat terkecil karena terjadinya masalah heteroskedastisitas pada model regresi dari distribusi *Gumbel*. Model regresi diperoleh dari transformasi logaritma dari fungsi distribusi kumulatif, dimana penentuan fungsi distribusi kumulatifnya menggunakan estimator *mean rank*. Metodologi ini diterapkan pada data curah hujan tahunan di Telangana, India Selatan dari tahun 1973-2010 yang bersumber dari penelitian Umarfarooque (2011). Hasil estimasi menunjukkan bahwa nilai parameter lokasi sebesar 507,204 mm dan parameter skala sebesar 423,729 mm.

Kata Kunci : Distribusi *Gumbel*, Estimator *Mean Rank*, Heteroskedastisitas, Kuadrat Terkecil Terboboti, Regresi *Rank*.

ABSTRACT

Least squares method based on the relationship between the empirical cumulative distribution also called rank regression, can be used to estimate the parameters of distribution. In this research, a weighted least squares method is used to estimate the parameters of the Gumbel distribution, where the weights are proportional to the inverse of the large sample variances of a function of the order statistics. This method is used as an alternative to the least squares method for the problem of heteroscedasticity in regression model of the Gumbel distribution. The regression model obtained from transformation logarithm of the cumulative distribution function, where the determination of the cumulative distribution function using the mean rank estimator. This methodology is applied to the data of annual rainfall (1973-2010) in Southern Telangana, India was sourced from research Umarfarooque (2011). The estimation results show that the value of the location parameter is 507,204 mm and scale parameter is 423,729 mm**.**

Keywords : Gumbel Distribution, Mean Rank Estimator, Heteroskedasticity, Weighted Least Squares, Rank Regression.

[*] *Prodi Statistika, Jurusan Matematika, Univesitas Hasanuddin*
email : anna.statistik@gmail.com

1. Pendahuluan

Statistika inferensia merupakan teknik pengambilan kesimpulan berdasarkan data yang diperoleh dari sampel untuk menggambarkan karakteristik atau ciri dari suatu populasi. Statistika inferensia meliputi dua hal penting, yaitu estimasi parameter dan pengujian hipotesis. Jika parameter populasi diketahui maka dilakukan pengujian hipotesis untuk menguji kebenaran dari asusmsi tentang parameter, tapi jika parameter populasi tidak diketahui maka dilakukan estimasi parameter.

Dalam analisis frekuensi data hidrologi pada data curah hujan, sangat jarang dijumpai seri data yang sesuai dengan distribusi normal. Salah satu distribusi yang sering digunakan adalah distribusi Gumbel (Arwin, 2007). Distribusi Gumbel diperkenalkan pertama kali oleh seorang ahli matematika Jerman Emil Gumbel (1891-1966). Fokus Gumbel adalah terutama pada aplikasi dari teori nilai ekstrim untuk masalah rekayasa, dalam pemodelan tertentu fenomena meteorologi seperti arus banjir tahunan. Menurut Waliesta (1997), distribusi Gumbel disebut juga distribusi nilai ekstrim tipe I, banyak digunakan untuk menyatakan kejadian debit tahunan (Arwin, 2007).

Berbagai metode telah diusulkan untuk mengestimasi parameter dari distribusi Gumbel. Salah satu metode yang biasa digunakan adalah metode kuadrat terkecil biasa. Metode ini dilakukan dengan mentransformasi distribusi Gumbel ke dalam model regresi rank yang berkaitan dengan statistik terurut. Secara matematika, estimasi parameter regresi ini dengan cara meminimumkan jumlah kuadrat *error*. Namun, salah satu asumsi pada regresi linier sederhana adalah variansi dari *error*-nya konstan atau homoskedastisitas (Allen, 1997). Penyimpangan terhadap asumsi homoskedastisitas sering terjadi. Dengan adanya pelanggaran asumsi ini, penggunaan metode kuadrat terkecil biasa memiliki efisiensi yang rendah sehingga perlu dilakukan tindakan perbaikan.

Menurut Zyl dan Schall (2011), statistik terurut $X_1 \leq \cdots \leq X_n$ tidak memiliki variansi konstan, sehingga model regresi rank yang terbentuk dari distribusi Gumbel adalah heteroskedastisitas dan metode kuadrat terkecil biasa tidak dapat digunakan. Untuk memaksimalkan efisiensi estimasi parameter, digunakan metode kuadrat terkecil terboboti dengan memberikan jumlah bobot yang tepat pada tiap titik data.

Pada tulisan ini, parameter distribusi Gumbel akan diestimasi dengan menggunakan metode kuadrat terkecil terboboti dengan menurunkan pendekatan bobot untuk variansi sampel yang besar dari sebuah fungsi dari statistik terurut untuk menstabilkan variansi dan diaplikasikan pada data curah hujan tahunan. Fungsi Distribusi Kumulatif (FDK) nantinya diestimasi menggunakan estimator *mean rank*.

2. Tinjauan Pustaka

2.1 Distribusi Gumbel

Distribusi Gumbel adalah suatu rumusan distribusi statistik. Distribusi Gumbel termasuk jenis distribusi nilai ekstrim. Fungsi distribusi kumulatif (FDK) dan fungsi kepadatan peluang (fkp) dari distribusi Gumbel adalah :

$$F(x) = \exp\left\{-\exp\left[-\frac{(x-\alpha)}{\beta}\right]\right\} \tag{1}$$

$$f(x) = \frac{1}{\beta}\exp\left\{\left[-\frac{(x-\alpha)}{\beta}\right] - \exp\left[-\frac{(x-\alpha)}{\beta}\right]\right\},\ -\infty \leq x \leq \infty, \beta > 0 \tag{2}$$

dimana :

$F(x)$ = FDK dari distribusi Gumbel
$f(x)$ = fkp dari distribusi Gumbel
α = parameter lokasi
β = parameter skala
x = variabel acak kontinu (Anggun Haryanto, 2011).

2.2 Regresi Linier Sederhana

Menurut Sembiring (1995), model regresi adalah model yang memberikan gambaran mengenai hubungan antara variabel bebas X dan variabel tidak bebas Y yang dipengaruhi oleh beberapa parameter regresi yang belum diketahui nilainya. Jika analisis regresi dilakukan untuk satu variabel bebas dengan satu variabel tidak bebas, maka regresi ini dinamakan regresi linear sederhana dengan model sebagai berikut :

$$Y_i = \theta_0 + \theta_1 X_i + e_i,\ i = 1, 2, \dots, n \tag{3}$$

dimana :

Y = variabel tidak bebas
θ_0, θ_1 = koefisien regresi
X = variabel bebas
e_i = *error* dengan asumsi e_i berdistribusi normal dengan rata-rata 0, variansi σ^2, dan saling independen (Kusuma, 2007).

Model regresi linear sederhana dalam notasi matriks sebagai berikut :

$$\boldsymbol{Y} = \boldsymbol{X\theta} + \boldsymbol{e}.$$

2.3 Metode Kuadrat Terkecil Biasa

Konsep dari metode kuadrat terkecil biasa adalah mengestimasi parameter regresi dengan memininumkan jumlah kuadrat *error*. Untuk memperoleh estimator pada persamaan (3) maka dilakukan dengan metode kuadrat terkecil biasa, yaitu :

$$\begin{aligned}
\text{SSR} &= \textstyle\sum_{i=1}^{n} e_i{}^2 \\
&= e_1{}^2 + e_2{}^2 + \cdots + e_n{}^2 \\
&= \begin{bmatrix} e_1 & e_2 & \dots & e_n \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \\
&= \boldsymbol{e}^t \boldsymbol{e} \\
&= (\boldsymbol{Y} - \boldsymbol{X\theta})^t (\boldsymbol{Y} - \boldsymbol{X\theta}),
\end{aligned} \tag{4}$$

dengan menurunkan persamaan (4) terhadap $\boldsymbol{\theta}$ dan menyamakan hasil turunannya terhadap nol, maka diperoleh estimasi untuk $\boldsymbol{\theta}$:

$$\widehat{\boldsymbol{\theta}} = (\boldsymbol{X^t X})^{-1} \boldsymbol{X^t Y} \tag{5}$$

(Abdul Aziz, 2007).

2.4 Metode Kuadrat Terkecil Terboboti

Salah satu asumsi yang harus dipenuhi agar model bersifat BLUE (*Best Linear Unbiased Estimator*) adalah harus terdapat variansi yang sama dari setiap *error*-nya atau homoskedastisitas, dengan kata lain $\text{var}(e_i) = \sigma^2$. Apabila asumsi ini tidak terpenuhi maka yang terjadi adalah sebaliknya, yakni heteroskedastisitas. Heteroskedastisitas berarti variansi *error* berbeda dari suatu observasi ke observasi lainnya (Firdaus, 2004).

Menurut Nachrowi dan Hardius (2006 alternatif model estimasi yang baik untuk berhadapan dengan heteroskedastisitas adalah metode kuadrat terkecil terboboti, karena disamping kuadrat terkecil terboboti memiliki kemampuan untuk menetralisir akibat dari pelanggaran asumsi heteroskedastisitas, kuadrat terkecil terboboti juga tidak kehilangan sifat ketakbiasan dan konsistensi dari model estimasi kuadrat terkecil terboboti.

Misalkan model yang digunakan adalah model pada persamaan (3), yaitu $\boldsymbol{Y} = \boldsymbol{X\theta} + \boldsymbol{e}$, dengan asumsi $\text{E}(\boldsymbol{e}) = \boldsymbol{0}$ dan $\text{var}(\boldsymbol{e}) = \text{E}(\boldsymbol{ee^t}) = \boldsymbol{V}\sigma^2$ dimana $\boldsymbol{V}$ adalah matriks definit positif berukuran $n \times n$ yang diketahui.

Jika $\boldsymbol{V}$ adalah matriks definit positif maka terdapat suatu matriks simetris $\boldsymbol{P}$ berukuran $n \times n$ yang nonsingular sedemikian sehingga $\boldsymbol{P^t P} = \boldsymbol{V}^{-1}$ dan $\boldsymbol{PVP^t} = \boldsymbol{I}$ (Judge dkk., 1988).

Misalkan bahwa $\boldsymbol{f} = \boldsymbol{Pe}$ sehingga $\mathrm{E}(\boldsymbol{f}) = \boldsymbol{0}$ dan $\mathrm{var}(\boldsymbol{f}) = \mathrm{E}(\boldsymbol{f}\boldsymbol{f}^{\boldsymbol{t}}) = \boldsymbol{I}\sigma^2$. Jika persamaan $\boldsymbol{Y} = \boldsymbol{X\theta} + \boldsymbol{e}$ digandakan dengan $\boldsymbol{P}$**,** maka akan diperoleh sebuah model baru sebagai berikut :

$$\boldsymbol{PY} = \boldsymbol{PX\theta} + \boldsymbol{Pe}$$

atau

$$\boldsymbol{Z} = \boldsymbol{Q\theta} + \boldsymbol{f}, \tag{6}$$

dengan $\boldsymbol{Z} = \boldsymbol{PY}, \boldsymbol{Q} = \boldsymbol{PX}$ dan $\boldsymbol{f} = \boldsymbol{Pe}$**.**

(Drapper and Smith, 1992).

Peminimalan $\boldsymbol{f}\boldsymbol{f}^{\boldsymbol{t}}$ terhadap $\boldsymbol{\theta}$ dengan cara sama pada metode kuadrat terkecil biasa untuk $\boldsymbol{\theta}$ pada model yang telah ditransformasikan adalah :

$$\begin{aligned}\widehat{\boldsymbol{\theta}} &= (\boldsymbol{Q}^{\boldsymbol{t}}\boldsymbol{Q})^{-\boldsymbol{1}}\boldsymbol{Q}^{\boldsymbol{t}}\boldsymbol{Z} \\ &= \left(\boldsymbol{X}^{\boldsymbol{t}}\boldsymbol{V}^{-\boldsymbol{1}}\boldsymbol{X}\right)^{-\boldsymbol{1}}\boldsymbol{X}^{\boldsymbol{t}}\boldsymbol{V}^{-\boldsymbol{1}}\boldsymbol{Y}.\end{aligned} \tag{7}$$

2.5 Statistik Terurut

Definisi 1 (Hogg dan Craig, 1995)

Misalkan $X_1, X_2, \dots, X_n$ merupakan sebuah sampel acak berukuran *n* dari sebuah distribusi kontinu yang mempunyai fkp $f(x)$ positif, untuk $a < x < b$. Jika $X_{(1)}$ adalah nilai terkecil dari $(X_1, X_2, \dots, X_n)$, $X_{(2)}$ adalah nilai terkecil kedua dari $(X_1, X_2, \dots, X_n)$,..., $X_{(n)}$ adalah nilai terbesar dari $(X_1, X_2, \dots, X_n)$. Maka akan berlaku hubungan sebagai berikut:

$$X_{(1)} < X_{(2)} < \dots < X_{(n)}.$$

Dalam hal ini, $X_{(i)}, i = 1,2, \dots, n$ dinamakan statistik terurut ke-*i* dari sampel acak $X_1, X_2, X_3, \dots, X_n$.

Teorema 1 (Casella dan Beger, 1990)

Misalkan $X_1, X_2, X_3, \dots, X_n$ merupakan sebuah sampel acak berukuran *n* dari sebuah distribusi populasi kontinu yang mempunyai fungsi kepadatan peluang $f(x)$ dan fungsi distribusi kumulatif $F(x)$. Maka, fungsi kepadatan peluang dari statistik terurut ke-*i* diberikan oleh :

$$\begin{aligned}g_i(x_{(i)}) &= \frac{n!}{(i-1)!(n-i)!} f(x_{(i)})\left[F(x_{(i)})\right]^{i-1}\left[1 - F(x_{(i)})\right]^{n-i}, \\ &\qquad\qquad\qquad\qquad a < x_{(i)} < b \\ &= 0, \text{lainnya.}\end{aligned} \tag{8}$$

2.6 Pendekatan Bobot untuk Variansi Sampel yang Besar

Bobot untuk regresi adalah invers dari variansi yang mendekati fungsi $\boldsymbol{\Lambda}$ skalar dari statistik terurut. Hal ini diasumsikan bahwa turunan dari $\boldsymbol{\Lambda}$ kontinu pada nilai harapan dari statistik terurut.

Misalkan $X_1, \dots, X_n$ menunjukkan suatu sampel berukuran *n* dari distribusi F dengan statistik terurut $X_{(1)} \leq \dots \leq X_{(n)}$ yang bersesuaian. Ekspresi kuadrat terkecil terboboti dengan memperhatikan parameter untuk meminimalkan $\sum_{i=1}^{n} w_i \left[\mathrm{E}\left(\boldsymbol{\Lambda}(x_{(i)})\right) - \boldsymbol{\Lambda}(x_{(i)})\right]^2$, dimana bobot untuk kuadrat *error* $e_i^2 = [\boldsymbol{\Lambda}(X_i) - \boldsymbol{\Lambda}(x_{(i)})]^2$ ke-*i* adalah $w_i = 1/\mathrm{var}\left(\boldsymbol{\Lambda}(x_{(i)}\right), i = 1, \dots, n$. Fungsi $\boldsymbol{\Lambda}$ tidak perlu menjadi fungsi linear dari statistik terurut.

Statistik $F\left(x_{(i)}\right), \dots, F\left(x_{(n)}\right)$ adalah distribusi beta dengan $F\left(x_{(i)}\right) \sim Beta(i, n - i + 1)$, dimana :

$$\begin{aligned}\mathrm{E}\left(F\left(x_{(i)}\right)\right) &= m_i = \frac{i}{n+1} \\ \mathrm{var}\left(F\left(x_{(i)}\right)\right) &= \frac{i(n-i+1)}{(n+2)(n+1)^2} = \frac{m_i(1-m_i)}{n+2}\end{aligned} \tag{9}$$

Misalkan X_i sedemikian sehingga $F^{-1}(X_i) = \frac{i}{n+1}$. Secara asimtotik diperoleh $\sqrt{n}[x_{(i)} - X_i] \xrightarrow{d} Z$, dimana $Z \sim N(0, \sigma_i^2)$ dengan $\sigma_i^2 = \frac{m_i(1-m_i)}{(n+2)(F'(X_i))^2}, i = 1, \dots, n$ dan $F'(m_i) = f(m_i)$ ada (Dasgupta, 2008).

Metode delta dapat diterapkan untuk mendapatkan perkiraan variansi skalar fungsi bernilai $\mathbf{\Lambda}$ dari statistik terurut, di mana diasumsikan bahwa turunan pertama dari $\mathbf{\Lambda}$ adalah kontinu pada X_i dan $\mathbf{\Lambda}'(X_i) \neq 0$. Selanjutnya,

$$\sqrt{n}(\mathbf{\Lambda}(x_{(i)}) - \mathbf{\Lambda}(X_i)) \sim N\left(0, \mathrm{var}(x_{(i)}) \left(\frac{d\mathbf{\Lambda}(x_{(i)})}{dx_{(i)}}\right)^2_{x_{(i)}=X_i}\right), i = 1, \dots, n$$

sehingga dapat disimpulkan bahwa

$$\mathrm{var}(\mathbf{\Lambda}(x_{(i)})) \approx \frac{m_i(1-m_i)}{(n+2)(f(X_i))^2} \left(\frac{d\mathbf{\Lambda}(x_{(i)})}{dx_{(i)}}\right)^2_{x_{(i)}=X_i}.$$

Selanjutnya jika $\mathbf{\Lambda}(x_{(i)})$ dalam bentuk $\mathbf{\Lambda}(x_{(i)}) = \mathbf{\Lambda}(F(x_{(i)}))$ dapat dilihat bahwa

$$\mathrm{var}(\mathbf{\Lambda}(F(x_{(i)})) \approx \frac{m_i(1-m_i)}{(n+2)(f(X_i))^2} \left(\left(\frac{d\mathbf{\Lambda}(\mathrm{F}(x_{(i)}))}{d\mathrm{F}(x_{(i)})}\right)\left(\frac{d(\mathrm{F}(x_{(i)}))}{d(x_{(i)})}\right)\right)^2_{x_{(i)}=X_i}$$

$$= \frac{m_i(1-m_i)}{(n+2)} \left(\frac{d\mathbf{\Lambda}(F(x_{(i)}))}{dF(x_{(i)})}\right)^2_{x_{(i)}=X_i} \tag{10}$$

(Zyl dan Schall, 2011).

2.7 Metode Regresi Rank Kuadrat Terkecil Terboboti

Metode kuadrat terkecil **y**ang didasarkan pada hubungan antara fungsi distribusi kumulatif empiris dan statistik terurut sebagai variabel bebas disebut juga regresi *rank*, dapat digunakan untuk mengestimasi parameter dari distribusi (Zyl dan Schall, 2011).

Metode ini dilakukan dengan mentransformasi fungsi distribusi kumulatif (FDK) dari suatu distribusi ke dalam bentuk linear. Selanjutnya dibentuk model regresi rank yang berkaitan dengan statistik terurut.

Estimasi parameter dari suatu distribusi diperoleh dengan menerapkan metode kuadrat terkecil terboboti pada persamaan (7) dengan bobot adalah invers dari variansi fungsi statistik terurut.

3. Hasil dan Pembahasan

3.1 Estimasi Parameter Distribusi Gumbel menggunakan Metode Regresi Rank Kuadrat Terkecil Terboboti

3.1.1 Transformasi Logaritma Fungsi Distribusi Kumulatif Distribusi Gumbel

Fungsi distribusi kumulatif Gumbel merupakan fungsi non-linear yang akan ditransformasi ke fungsi linear dengan menggunakan transformasi logaritma yaitu :

$$F(x) = \exp\left(-\exp\left(-\frac{(x-\alpha)}{\beta}\right)\right)$$

$$-\log(-\log F(x)) = \frac{(x-\alpha)}{\beta}. \tag{11}$$

3.1.2 Model Regresi Rank

Misalkan $x_1, x_2, \dots, x_n$ adalah sampel acak berukuran *n* dengan statistik terurut $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ dengan asumsi bahwa sampel berdistribusi Gumbel dua parameter dengan parameter α dan β tidak diketahui, dimana $\hat{\alpha}$ dan $\hat{\beta}$ menyatakan estimator dari α dan β.

Untuk sampel berukuran *n* dari distribusi Gumbel dengan statistik terurut $x_{(1)}, x_{(2)}, \dots, x_{(n)}$, persamaan (11) menjadi :

$$-\log(-\log \hat{F}(x_{(i)})) = \frac{1}{\beta}x_{(i)} - \frac{\alpha}{\beta} \tag{12}$$

dimana $i = 1, 2, \ldots, n$ adalah nomor urut ke-i dan $\hat{F}(x_{(i)})$ adalah estimator nonparametrik dari $F(x_{(i)})$, yaitu estimator *mean rank* $\hat{F}(x_{(i)}) = m_i = \frac{i}{n+1}$.

Dari persamaan (12) dapat ditentukan model regesi sebagai berikut :

$$-\log(-\log \hat{F}(x_{(i)})) = \frac{1}{\beta}x_{(i)} - \frac{\alpha}{\beta} + e_i. \tag{13}$$

Misalkan bahwa

$$Y_i = -\log(-\log \hat{F}(x_{(i)}))$$
$$X_i = x_{(i)}$$
$$\theta_0 = -\frac{\alpha}{\beta}$$
$$\theta_1 = \frac{1}{\beta},$$

sehingga persamaan (13) dapat ditulis sebagai berikut :

$$Y_i = \theta_0 + \theta_1 X_i + e_i.$$

3.1.3 Estimasi Parameter

Menurut Zyl dan Schall (2011), statistik urut $x_{(1)} \leq \cdots \leq x_{(n)}$ tidak memiliki variansi konstan, sehingga model regresi yang terbentuk dari distribusi Gumbel pada persamaan (13) adalah heteroskedastisitas dan metode kuadrat terkecil tidak dapat digunakan melainkan perlu dilakukan tindakan perbaikan.

Metode yang digunakan untuk mengestimasi parameter distribusi Gumbel adalah metode regresi rank kuadrat terkecil terboboti.

3.1.3.1 Penentuan Nilai Pembobot $\boldsymbol{w_i}$

Misalkan $x_1, \ldots, x_n$ menunjukkan suatu sampel berukuran n dari distribusi $F(x) = F(x; \alpha, \beta)$ dimana α dan β adalah parameter yang tidak diketahui dengan statistik terurut $x_{(1)} \leq \cdots \leq x_{(n)}$ yang bersesuaian. Metode kuadrat terkecil terboboti digunakan untuk mengestimasi parameter dengan cara meminimalkan $\sum_{i=1}^{n} w_i\, [\boldsymbol{\Lambda}(\hat{F}(x_{(i)})) - \boldsymbol{\Lambda}(F(x_{(i)}))]^2$, dimana bobot untuk kuadrat *error* $e_i^2 = [\boldsymbol{\Lambda}(\hat{F}(x_{(i)})) - \boldsymbol{\Lambda}(F(x_{(i)}))]^2$ ke-i adalah $w_i = \frac{1}{\mathrm{var}\,(e_i)} = \frac{1}{\mathrm{var}\,(\boldsymbol{\Lambda}(F(x_{(i)})))}, i = 1, \ldots, n.$

Misalkan sebuah sampel berukuran n dari distribusi Gumbel dua parameter dengan parameter α dan β. Pendekatan variansi dari $\left(-\log\left(-\log \hat{F}\left(x_{(i)}\right)\right)\right)$ adalah :

$$\mathrm{var}\left(-\log\left(-\log \hat{F}\left(x_{(i)}\right)\right)\right) \approx \frac{m_i(1-m_i)}{(n+2)}\left(\frac{d\left(-\log\left(-\log \hat{F}\left(x_{(i)}\right)\right)\right)}{d\hat{F}\left(x_{(i)}\right)}\right)^2$$

$$\mathrm{var}\left(-\log\left(-\log \hat{F}\left(x_{(i)}\right)\right)\right) = \frac{\left(1-\frac{i}{n+1}\right)}{(n+2)\left(\frac{i}{n+1}\right)\left(\log\left(\frac{i}{n+1}\right)\right)^2}$$

Sehingga diperoleh pembobot :

$$w_i = \frac{1}{\mathrm{var}\left(-\log\left(-\log \hat{F}\left(x_{(i)}\right)\right)\right)}$$

$$= \frac{(n+2)\left(\frac{i}{n+1}\right)\left(\log\left(\frac{i}{n+1}\right)\right)^2}{\left(1-\frac{i}{n+1}\right)}, i = 1,2, \ldots, n. \tag{14}$$

3.1.3.2 Penentuan Nilai Estimasi Parameter α dan β dari Distribusi Gumbel

Untuk mengestimasi parameter pada persamaan (13) yang bersifat heteroskedastisitas, digunakan metode kuadrat terkecil terboboti. Persamaan (13) dapat dibentuk ke dalam model regresi linear sederhana, yaitu :

$$Y_i = \theta_0 + \theta_1 X_i + e_i.$$

Model regresi linear sederhana dalam notasi matriks sebagai berikut :

$$\boldsymbol{Y} = \boldsymbol{X\theta} + \boldsymbol{e},$$

Untuk menentukan estimasi θ_0 dan θ_1 yang kemudian dinotasikan dengan $\hat{\theta}_0$ dan $\hat{\theta}_1$, dilakukan dengan menerapkan persamaan (7) yaitu :

$$\hat{\boldsymbol{\theta}} = \left(\boldsymbol{X^t V^{-1} X}\right)^{-1} \boldsymbol{X^t V^{-1} Y}$$

$$= \begin{pmatrix} \frac{\sum_{i=1}^{n} w_i X_i{}^2 \sum_{i=1}^{n} w_i Y_i - \sum_{i=1}^{n} w_i X_i \sum_{i=1}^{n} w_i X_i Y_i}{\sum_{i=1}^{n} w_i \sum_{i=1}^{n} w_i X_i{}^2 - \left(\sum_{i=1}^{n} w_i X_i\right)^2} \\ \frac{\sum_{i=1}^{n} w_i \sum_{i=1}^{n} w_i X_i Y_i - \sum_{i=1}^{n} w_i X_i \sum_{i=1}^{n} w_i Y_i}{\sum_{i=1}^{n} w_i \sum_{i=1}^{n} w_i X_i{}^2 - \left(\sum_{i=1}^{n} w_i X_i\right)^2} \end{pmatrix}.$$

Dengan mengganti nilai $Y_i = -\log\left(-\log\left(\hat{F}\left(x_{(i)}\right)\right)\right)$ dan $X_i = x_{(i)}$, diperoleh

$$\hat{\theta}_0 = \frac{\sum_{i=1}^{n} w_i x_{(i)}{}^2 \sum_{i=1}^{n} w_i\left(-\log\left(-\log\left(\hat{F}\left(x_{(i)}\right)\right)\right)\right) - \sum_{i=1}^{n} w_i x_{(i)} \sum_{i=1}^{n} w_i x_{(i)}\left(-\log\left(-\log\left(\hat{F}\left(x_{(i)}\right)\right)\right)\right)}{\sum_{i=1}^{n} w_i \sum_{i=1}^{n} w_i x_{(i)}{}^2 - \left(\sum_{i=1}^{n} w_i x_{(i)}\right)^2} \tag{15}$$

dan

$$\hat{\theta}_1 = \frac{\sum_{i=1}^{n} w_i \sum_{i=1}^{n} w_i x_{(i)}\left(-\log\left(-\log\left(\hat{F}\left(x_{(i)}\right)\right)\right)\right) - \sum_{i=1}^{n} w_i x_{(i)} \sum_{i=1}^{n} w_i\left(-\log\left(-\log\left(\hat{F}\left(x_{(i)}\right)\right)\right)\right)}{\sum_{i=1}^{n} w_i \sum_{i=1}^{n} w_i x_{(i)}{}^2 - \left(\sum_{i=1}^{n} w_i x_{(i)}\right)^2} \tag{16}$$

Nilai estimasi parameter $\hat{\alpha}$ dan $\hat{\beta}$ dari distribusi Gumbel, dipersoleh dari :

$$\begin{pmatrix} \hat{\theta}_0 \\ \hat{\theta}_1 \end{pmatrix} = \begin{pmatrix} -\hat{\alpha}/\hat{\beta} \\ 1/\hat{\beta} \end{pmatrix}.$$

Sehingga

$$\hat{\beta} = \frac{\sum_{i=1}^{n} w_i \sum_{i=1}^{n} w_i x_{(i)}{}^2 - \left(\sum_{i=1}^{n} w_i x_{(i)}\right)^2}{\sum_{i=1}^{n} w_i \sum_{i=1}^{n} w_i x_{(i)}\left(-\log\left(-\log\left(\hat{F}\left(x_{(i)}\right)\right)\right)\right) - \sum_{i=1}^{n} w_i x_{(i)} \sum_{i=1}^{n} w_i\left(-\log\left(-\log\left(\hat{F}\left(x_{(i)}\right)\right)\right)\right)} \tag{17}$$

dan

$$\hat{\alpha} = -\frac{\sum_{i=1}^{n} w_i x_{(i)}{}^2 \sum_{i=1}^{n} w_i\left(-\log\left(-\log\left(\hat{F}\left(x_{(i)}\right)\right)\right)\right) - \sum_{i=1}^{n} w_i x_{(i)} \sum_{i=1}^{n} w_i x_{(i)}\left(-\log\left(-\log\left(\hat{F}\left(x_{(i)}\right)\right)\right)\right)}{\sum_{i=1}^{n} w_i \sum_{i=1}^{n} w_i x_{(i)}\left(-\log\left(-\log\left(\hat{F}\left(x_{(i)}\right)\right)\right)\right) - \sum_{i=1}^{n} w_i x_{(i)} \sum_{i=1}^{n} w_i\left(-\log\left(-\log\left(\hat{F}\left(x_{(i)}\right)\right)\right)\right)}. \tag{18}$$

3.2 Aplikasi pada Data

Data yang digunakan dalam penelitian ini merupakan data sekunder berupa data curah hujan tahunan di Telangana, India Selatan dari tahun 1973 – 2010 yang diperoleh dari *Indian Journal of Natural Sciences*.

Variabel yang digunakan dalam penelitian ini sebagai berikut :

X : data curah hujan tahunan (mm)

Y : transformasi logaritma dari fungsi distribusi kumulatif distribusi Gumbel, dengan jumlah sampel $n = 38$.

3.2.1 Uji Kesesuaian Distribusi

Untuk menguji apakah data curah hujan tahunan di Telangana, India Selatan dari tahun 1973 – 2010 mengikuti distribusi Gumbel, digunakan uji *Chi-square* dengan hasil sebagai berikut :

1. Hipotesis

   H_0 : data mengikuti distribusi Gumbel

   H_1 : data tidak mengikuti distribusi Gumbel
2. Tingkat signifikansi sebesar 5% $= 0.05$
3. Kriteria keputusan

   Tolak H_0 jika χ^2_h (nilai statistik)> $\chi^2_{\alpha^*,df}$.
4. Kesimpulan

Hasil uji *Chi-Square* dengan *software Easy Fit* 5.5

Tabel 1. Hasil Uji *Chi-Square*

Chi-Square					
Derajat kebebasan Nilai statistik (χ_h^2)	4 0.06235				
Tingkat signifikansi (α^*)	0.2	0.1	0.05	0.02	0.01
Nilai kritis	5.9886	7.7794	9.4877	11.668	13.227

Sumber : Hasil olah data, 2014.

Berdasarkan output *software Easy Fit* 5.5 pada Tabel 1 diperoleh nilai kritis untuk tingkat signifikansi $\alpha^* = 0.05$ dan derajat kebebasan $df = 4$ adalah 9.4877 dan nilai statistik adalah 0.06235. Karena nilai statistik $(\chi_h^2) <$ nilai kritis $(\chi^2_{\alpha^*,df})$ yaitu $0.06235 < 9.4877$, maka H_0 diterima. Oleh karena itu, dapat disimpulkan bahwa data curah hujan tahunan di Telangana, India Selatan dari tahun 1973 – 2010 berdistribusi Gumbel.

3.2.2 Pendeteksian Heteroskedastisitas

Untuk mendeteksi adanya heteroskedastistas digunakan uji *White*. Untuk menunjukkan apakah data curah hujan tahunan di Telangana, India Selatan dari tahun 1973 – 2010 mengalami gangguan heteroskedastisitas digunakan bantuan *software* SPSS 17 dengan hasil sebagai berikut :

1. Formulasi hipotesis
   H_0 : tidak terdapat masalah heteroskedastisitas dalam model.
   H_1 : terdapat masalah heteroskedastisitas dalam model.
2. Tingkat signifikansi sebesar 5% $= 0.05$
3. Kriteria pengujian
   Tolak H_0 jika $n \times R^2 > \chi^2_{\alpha^*,df}$, dimana $\chi^2_{\alpha^*,df}$ adalah tabel distribusi *Chi-square* pada tingkat signifikansi $\alpha^* = 0.05$ dan deraja kebebasan $df = 2$.
4. Kesimpulan
   Hasil uji *White* dengan *software SPSS* 17.

Tabel 2. Hasil Uji *White*

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. *Error* of the Estimate
1	.954[a]	.909	.904	.01242

a. Predictors: (Constant), x2, X

b. Dependent Variable: e2

Sumber : Hasil olah data, 2014

Output untuk uji heteroskedastisitas pada Tabel 2 memberikan hasil perkalian banyak observasi dengan koefisien determinasi sebagai berikut :

$$nR^2 = 38 \times 0.909$$
$$= 34.542,$$

dipihak lain didapatkan nilai $\chi^2_{\alpha^*,df}$ tabel dengan derajat kebebasan $df = 2$ dengan tingkat kesalahan $\alpha^* = 0.05$ sebagai berikut :

$\chi^2_{0.05,2} = 5.991$ (dapat dilihat pada tabel distribusi Chi-Square)

dan didapatkan perbandingan sebagai berikut :

$$nR^2 > \chi^2_{\alpha^*,df},$$

dengan demikian H_0 ditolak dan dapat disimpulkan bahwa terdapat masalah heteroskedastisitas dalam model.

3.2.3 Nilai Estimasi Parameter Distribusi Gumbel dengan Metode Kuadrat Terkecil Terboboti

Dengan menggunakan persamaan (15) dan (16) diperoleh nilai $\hat{\theta}_0$ dan $\hat{\theta}_1$ sebagai berikut :

$\hat{\theta}_0 = -1.197$ dan $\hat{\theta}_1 = 0.00236$,

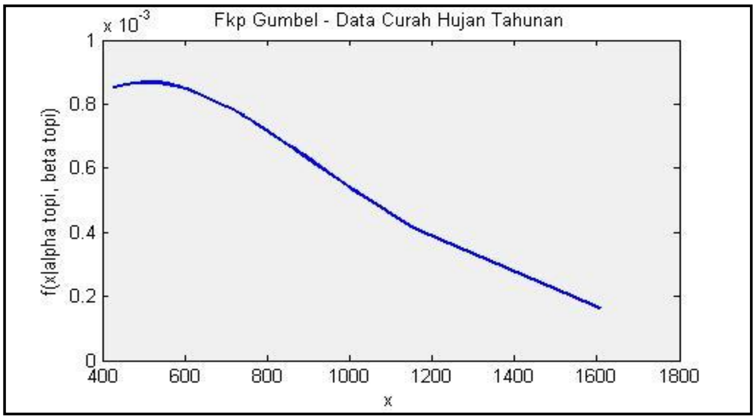
sehingga diperoleh nilai $\hat{\beta} = 423.729$ dan $\hat{\alpha} = 507.204$.

Dari hasil pengolahan data curah hujan tahunan di Telangana, India Selatan dari tahun (1973-2010) diperoleh nilai estimator parameter distribusi Gumbel adalah $\hat{\alpha} = 507.204$ dan $\hat{\beta} = 423.729$. Sehingga, model distribusi Gumbel-nya adalah

$$f(x_i) = \frac{1}{423.729} \exp\left\{\left[-\frac{(x_i - 507.204)}{423.729}\right] - exp\left[-\frac{(x_i - 507.204)}{423.729}\right]\right\},$$
$$-\infty \leq x \leq \infty$$

dimana x_i menyatakan data curah hujan tahunan tahun ke-*i* di Telangana, India Selatan.

3.2.4 Plot Data Distribusi Gumbel



Gambar 1 Grafik Data Curah Hujan Tahunan di Telangana, India Selatan Terhadap Peluang Gumbel

Gambar 1 menunjukkan fungsi kepadatan peluang distribusi Gumbel pada data curah hujan tahunan di Telangana, India Selatan pada tahun 1973 – 2010 dengan nilai parameter lokasi dan skala masing - masing adalah $\hat{\alpha} = 507.204$ dan $\hat{\beta} = 423.729$. Parameter lokasi ($\hat{\alpha}$) menunjukkan titik pemusatan data. Semakin besar nilai parameter lokasi, maka distribusi data akan bergeser ke kanan, begitu pula sebaliknya, semakin kecil nilai parameter lokasi, maka distribusi data akan bergeser ke kiri. Sedangkan, parameter skala ($\hat{\beta}$) menunjukkan sebaran dari distribusi data. Semakin besar nilai parameter skala, maka distribusi data akan semakin menyebar, begitu pula sebaliknya, semakin kecil nilai parameter skala, maka distribusi data semakin menyempit.

4. Kesimpulan dan Saran

4.1 Kesimpulan

Dari hasil analisis yang telah dilakukan dan berdasarkan penjelasan yang telah diberikan, maka dapat diambil beberapa kesimpulan sebagai berikut:

1. Hasil estimasi parameter distribusi Gumbel menggunakan metode regresi rank kuadrat terkecil terboboti adalah :

$$\hat{\beta} = \frac{\sum_{i=1}^{n} w_i \sum_{i=1}^{n} w_i x_{(i)}{}^2 - \left(\sum_{i=1}^{n} w_i x_{(i)}\right)^2}{\sum_{i=1}^{n} w_i \sum_{i=1}^{n} w_i x_{(i)}\left(-\log\left(-\log\left(\hat{F}(x_{(i)})\right)\right)\right) - \sum_{i=1}^{n} w_i x_{(i)} \sum_{i=1}^{n} w_i\left(-\log\left(-\log\left(\hat{F}(x_{(i)})\right)\right)\right)}$$

dan

$$\hat{\alpha} = -\frac{\sum_{i=1}^{n} w_i x_{(i)}{}^2 \sum_{i=1}^{n} w_i\left(-\log\left(-\log\left(\hat{F}(x_{(i)})\right)\right)\right) - \sum_{i=1}^{n} w_i x_{(i)} \sum_{i=1}^{n} w_i x_{(i)}\left(-\log\left(-\log\left(\hat{F}(x_{(i)})\right)\right)\right)}{\sum_{i=1}^{n} w_i \sum_{i=1}^{n} w_i x_{(i)}\left(-\log\left(-\log\left(\hat{F}(x_{(i)})\right)\right)\right) - \sum_{i=1}^{n} w_i x_{(i)} \sum_{i=1}^{n} w_i\left(-\log\left(-\log\left(\hat{F}(x_{(i)})\right)\right)\right)},$$

dimana

$$w_i = \frac{(n+2)\left(\frac{i}{n+1}\right)\left(\log\left(\frac{i}{n+1}\right)\right)^2}{\left(1-\frac{i}{n+1}\right)}.$$

2. Berdasarkan hasil estimasi pada data curah hujan tahunan di Telangan, India Selatan dari tahun (1973-2010), maka diperoleh model distribusi Gumbel sebagai berikut :

$$f(x_i) = \frac{1}{423.729}\exp\left\{\left[-\frac{(x_i-507.204)}{423.729}\right] - exp\left[-\frac{(x_i-507.204)}{423.729}\right]\right\},\ -\infty \leq x \leq \infty.$$

Model tersebut menunjukkan bahwa nilai parameter lokasi sebesar 507,204 mm dan parameter skala sebesar 423,729 mm**.**

4.2 Saran

Penelitian lebih lanjut dapat menggunakan estimator *median rank* untuk penentuan distribusi kumulatifnya serta melakukan analisis perbandingan efisiensi relatif antara metode estimasi kuadrat terkecil terboboti dan metode maksimum likelihood atau metode estimasi lainnya untuk mengestimasi parameter distribusi Gumbel.

Daftar Pustaka

Agus, I. dan Hanwar, S. (2011). *Uji Kesesuaian Chi-Kuadrat Data Hujan Catchment Area Taratak Timbulun Kabupaten Pesisir Selatan.* Padang: Jurusan Teknik Sipil Politeknik Negeri Padang.

Allen, M.P. (1997). *Understanding Regression Analysis*. New York: Plenum Press.

Arwin. (2007). *Bahan Kuliah Hidrologi.* Bandung: Institut Teknologi Bandung.

Aziz, A. (2007). *Buku Ajar Ekonometrika Teori dan Analisis Matematis.* Malang: Jurusan Matematika.

Casella, G. dan Berger, L.R. (1990). *Statistical Inference*. California: Duxbury Press.

Dasgupta, A. (2008). *Asymptotic Theory of Statistics and Probability.* New York: Springer Texts in statistics.

Drapper, N. dan Smith, H. (1992). *Analisis Regresi Terapan* (Edisi Kedua). Terjemahan Oleh Bambang Sumantri. Jakarta: Gramedia Pustaka Utama.

Firdaus, M. (2004). *Ekonometrika Suatu Pendekatan Aplikatif*. Jakarta: Bumi Aksara.

Haryanto, A. (2011). *Penaksiran Parameter Model Nested Logit*. Depok: Universitas Indonesia.

Hogg, R. V. dan Craig, A. T. (1995). *Introduction to Mathematical Statistics, Fifth Edition.* New Jersey: Prentice Hall.

Johnston, J. (1972). *Econometric Methods*. 2d ed. New York: McGraw-Hill.

Judge, G. G., dkk. (1988). *Introduction to the Theory and Practice of Econometrics,* 2nd Ed. New York: John Wiley and Sons.

Kresna, J.A. (2014). *Distribusi Statistika Terurut, Bagian Kedua.* Makassar: Jurusan Matematika Unhas.

Kusuma. (2007). *Metode Rank Nonparametrik pada Model Regresi Linear*. Surakarta: Universitas Sebelas Maret.

Momin, U., dkk. (2011). *Rainfall Analysis for Crop Planning in Semi Arid Region of Southern Telangana, India*. Indian Journal Of Natural Sciences, pp. 450 – 458.

Nachrowi, D.N. dan Usman, H. (2006). *Pendekatan Populer dan Praktis Ekonometrika untuk Analisis Ekonomi dan Keuangan.* Jakarta: FEUI.

Reliability Hotwire, Reliability Basics. Overview of The Gumbel, Loglogistic, and Gamma Distribution. 2005. *http://www.weibull.com/hotwire/issue56/relbasics56.htm*, diakses tanggal 17 Juli 2014 pukul 9.22.

Zyl, V.M.J. dan Robert, S. (2011). *Parameter Estimation Through Weighted Least-Squares Rank Regression with Specific Reference to The Weibull and Gumbel Distributions.* Communications in Statistics – Simulation and Computation, 41:9, pp. 1654-1666.